THE LIMITING BEHAVIOR OF MULTIPLE ROOTS OF THE LIKELIHOOD EQUATION *

by

Michael D. Perlman

Technical Report No. 125

July, 1969

* This research was supported by National Science Foundation Grant GP-9593.

1. Introduction.

Let $X_1, X_2, \cdots$ be an infinite sequence of independent, identically distributed (i.i.d.) random variables (possibly vector-valued) with common density $p(x, \theta)$ with respect to some measure μ. This includes both absolutely continuous and discrete distributions. Here θ is a real-valued unknown parameter whose range we denote by Ω, a subset of the real line. It is assumed throughout that Ω is an interval (a, b), with $-\infty \le a < b \le +\infty$, which may be open, half-open, or closed. For a given sample outcome $(x_1, \cdots, x_n)$ the likelihood function (abbreviated LF) is

$$L_n(\theta) = L_n(\theta, x_1, \cdots, x_n) = \prod_{i=1}^{n} p(x_i, \theta)$$

and the likelihood equation (LEQ) is

$$\frac{\partial \log L_n(\theta, x_1, \cdots, x_n)}{\partial\theta} = 0. \tag{1.1}$$

Let $S(x_1, \cdots, x_n)$ denote the set (possibly empty) of all solutions of the LEQ. In this paper we study the behavior of the set $S(x_1, \cdots, x_n)$ as $n \to \infty$, with particular emphasis on the case of "multiple roots" where $S(x_1, \cdots, x_n)$ can contain more than one solution for some sample sequences.

Huzurbazar (1948) proved a result which is commonly described by the phrase "a consistent root of the LEQ is unique" (see Proposition 3 in Section 2). There are, however, ambiguities in the notions "consistent" and "root of the LEQ" which make this statement unclear. First, consistency is a limiting property of a sequence of estimators, say $\{T_n\}$, so for any $N < \infty$, T_n can be altered arbitrarily for all $n \le N$

without destroying consistency. In fact, if $\{T_n\}$ is a strongly consistent sequence of estimators of θ, i.e., $P[T_n \to \theta] = 1$, and if $\{T_n^*\}$ is another sequence such that $P[T_n^* = T_n$ for all sufficiently large $n] = 1$, then $\{T_n^*\}$ is also strongly consistent for θ. Next, by a "root of the LEQ" is meant any sequence of measurable functions $\{\hat{\theta}_n\} = \{\hat{\theta}_n(x_1, \cdots, x_n)\}$ such that for each n and each sample outcome $(x_1, \cdots, x_n)$, $\hat{\theta}_n(x_1, \cdots, x_n)$ is a solution of (1.1), i.e., $\hat{\theta}_n \in S(x_1, \cdots, x_n)$. If, in general, $S(x_1, \cdots, x_n)$ may contain more than one point, then there are infinitely many "roots of the LEQ" as here defined. (The Cauchy family of densities with an unknown location parameter illustrates this situation -- see Sections 5 and 6, and also Barnett (1966).) Thus, if $\{\hat{\theta}_n\}$ is a strongly consistent root of the LEQ and if $\{\hat{\theta}_n^*\}$ is another root such that $P[\hat{\theta}_n^* = \hat{\theta}_n$ for all sufficiently large $n] = 1$, then $\{\hat{\theta}_n^*\}$ is also a strongly consistent root. Therefore there may be many consistent roots of the LEQ, despite Huzurbazar's result.

In this paper we avoid these ambiguities by studying the behavior of the set $S(x_1, \cdots, x_n)$ as $n \to \infty$. In Section 2 we extend Huzurbazar's result by showing (Theorem 4) that with probability 1 all but one member of $S(x_1, \cdots, x_n)$ is bounded away from the true parameter value θ_0 as $n \to \infty$, and the exceptional member approaches θ_0 (assuming regularity conditions). In Section 3 we extend this result by showing that with probability 1, all members of $S(x_1, \cdots, x_n)$ except one approach the boundary of Ω if the Kullback-Leibler "distance" $I(\theta_0, \theta)$ increases as θ moves away from θ_0 (Theorem 8). If in addition

$(d/d\theta)I(\theta_0, \theta)$ is bounded away from 0 as θ approaches the boundary of Ω, then with probability 1, $S(x_1, \cdots, x_n)$ has exactly one member for all large n (Theorems 9, 11, 12). Thus, the maximum likelihood estimator (MLE) is strongly consistent in this case. The MLE, if it exists, is that point $\hat{\theta}_n$ such that $L_n(\hat{\theta}_n) = \sup \{L_n(\theta):\theta \in \Omega\}$. These results are applied to generalized monotone likelihood ratio families and location parameter families in Section 4. Some special cases are discussed in Section 5, including the Cauchy distribution and the bivariate normal distribution with known variances and unknown correlation. Section 6 contains some remarks concerning maximum likelihood estimation which were suggested by Barnett's (1966) study of the Cauchy distribution with unknown location parameter. Some details omitted from the discussion are presented in the Appendix.

Throughout this paper, θ_0 denotes the true (but unknown) parameter value, and the probability and expectation symbols P and E refer to the probability distribution under θ_0. The phrase "almost all x" refers to a set having probability 1 under θ_0. We assume that θ_0 is an interior point of Ω. For later convenience let X denote a random variable with density $p(x, \theta_0)$. For any $\delta > 0$ let I_δ denote the interval $[\theta_0-\delta, \theta_0+\delta]$.

Since we are concerned with the behavior of the LF $L_n(\theta)$ for $n \to \infty$, we must consider the probability space $(\mathfrak{X}, \mathfrak{B}, P)$ where $\mathfrak{X}$ is the set of all infinite sequences $(x_1, \cdots)$, $\mathfrak{B}$ is the usual (product) Borel field on $\mathfrak{X}$, and P is the product probability measure determined by the fact that $X_1, X_2, \cdots$ are i.i.d. with common density $p(x, \theta_0)$. For each n let A_n denote a subset of $\mathfrak{X}$ which depends only on the first n coordinates, i.e., $(x_1, \cdots, x_n, x_{n+1}, \cdots) \in A_n$ implies $(x_1, \cdots, x_n, x'_{n+1}, \cdots) \in A_n$ for all possible $x'_{n+1}, \cdots$. When we refer to the event (subset of $\mathfrak{X}$)

$$\{(x_1, \cdots, x_n) \text{ in } A_n \text{ for all sufficiently large } n\}$$

we mean the event

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

For each $(x_1, \cdots)$ in $\liminf A_n$, there exists an integer $N(x_1, \cdots)$ (depending on the sample sequence) such that $(x_1, \cdots, x_n) \in A_n$ for all $n \geq N(x_1, \cdots)$. The phrase "for all sufficiently large n" is abbreviated "fasln". We shall repeatedly use the fact that

$$(1.2) \qquad (\liminf A_n) \cap (\liminf B_n) = \liminf (A_n \cap B_n).$$

The term "relative maximum" is abbreviated throughout as "rel max".

2. Local limiting behavior of $S(x_1, \cdots, x_n)$.

In this section we discuss the behavior of the solutions of the LEQ in a neighborhood of θ_0. We need the following regularity conditions:

C1. If $\theta_1 \neq \theta_2$ then $p(x, \theta_1)$ and $p(x, \theta_2)$ determine distinct distributions.

C2. There is an open interval $I \subset \Omega$ containing θ_0 such that for almost all x, $p(x, \theta)$ is a continuous function of θ on I.

C3. For almost all x, $(\partial/\partial\theta) \log p(x, \theta)$ exists for all θ in I.

C4. For almost all x, $(\partial^2/\partial\theta^2) \log p(x, \theta)$ exists for all θ in I.

C5. The integral

$$I(\theta_0) \equiv -\int [\frac{\partial^2 \log p(x, \theta_0)}{\partial \theta^2}] \; p(x, \theta_0) d\mu(x)$$

$$= -E[\frac{\partial^2 \log p(X, \theta_0)}{\partial \theta^2}]$$

exists and $I(\theta_0) > 0$.

C6. There exists a nonnegative function $M(x)$ with $E_{\theta_0}[M(X)] < \infty$ and for all $\epsilon > 0$ there exists $\delta(\epsilon) > 0$, such that for almost all x

$$\sup_{|\theta-\theta_0|\le\delta(\epsilon)} \left|\frac{\partial^2 \log p(x, \theta)}{\partial \theta^2} - \frac{\partial^2 \log p(x, \theta_0)}{\partial \theta^2}\right| \le \epsilon M(x).$$

C7. In C6, replace "for all $\epsilon > 0$" by "for some $\epsilon < I(\theta_0)/E[M(X)]$".

Note that if the equation $\int p(x, \theta) d\mu(x) = 1$ can be twice differentiated under the integral sign at θ_0, then $I(\theta_0)$ is the Fisher Information Number; i.e.,

$$I(\theta_0) = \mathrm{Var}[(\partial/\partial\theta) \log p(X, \theta_0)] \tag{2.1}$$

$$= E[(\partial/\partial\theta) \log p(X, \theta_0)]^2.$$

The first proposition presented below is well know (see Cramer (1946) or Rao (1965)). Propositions 2 and 3 appear in Huzurbazar (1948).

Proposition 1. If C1 and C2 are true, then for all $\delta > 0$

$$P[\exists \text{ at least one rel max of } L_n(\theta) \text{ in } I_\delta \text{ fasln}] = 1.$$

If C3 also holds, then

$$P[\exists \text{ at least one solution of the LEQ in } I_\delta \text{ fasln}] = 1,$$

and this solution gives a relative maximum. Hence there exists at least one strongly consistent root of the LEQ.

Proposition 2. (Huzurbazar). Suppose C4, C5 and C6 are true. Let $\bar{\theta}_n = \bar{\theta}_n(x_1, \cdots, x_n)$ be any strongly consistent estimator of θ_0 (not necessarily a root of the LEQ). Then

$$P\left[\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2} \to -I(\theta_0)\right] = 1 \tag{2.2}$$

so that

$$P\left[\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2} < 0 \quad \text{fasln}\right] = 1. \tag{2.3}$$

If C6 is replaced by C7, (2.3) remains true.

Proposition 3. (Huzurbazar). If C4, C5, and C6 are true, if both $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are roots of the LEQ, and if both $\{\hat{\theta}_n\}$ and $\{\hat{\theta}_n^*\}$ are strongly consistent sequences of estimators for θ_0, then

$$P[\hat{\theta}_n = \hat{\theta}_n^* \quad \text{fasln}] = 1.$$

Remark: The conditions given here are weaker than those assumed by Huzurbazar. Our conditions will be used to prove Theorem 4, below, which is stronger than Proposition 3 and which implies (2.3). The result (2.2) is proved in Appendix I.

THEOREM 4. Suppose that C1, C4, C5, and C7 are true. Then there exists $\delta_0 > 0$ such that

$$P\left[\sup_{|\theta-\theta_0| \le \delta_0} \left(\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2}\right) < 0 \quad \text{fasln}\right] = 1.$$

Thus by Proposition 1, for any $0 < \delta \le \delta_0$

$$P[\exists \text{ exactly one solution of the LEQ in } I_\delta \quad \text{fasln}] = 1,$$

and this solution gives a relative maximum of the LF.

Proof. Since the Strong Law of Large Numbers (SLLN) implies

$$(2.4)\qquad \frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log p(X_i, \theta_0)}{\partial\theta^2} \to -I(\theta_0)$$

with probability 1, for any $\eta_1 > 0$ we have

$$P[\frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta^2} < -I(\theta_0)+\eta_1 \text{ fasln}] = 1.$$

By C7, there exist $\epsilon < I(\theta_0)/E[M(X)]$ and $\delta_0 = \delta(\epsilon) > 0$ such that

$$\sup_{|\theta-\theta_0|\le\delta_0} \left|\frac{1}{n}\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2} - \frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta^2}\right| \le \frac{\epsilon}{n}\sum_{i=1}^{n} M(x_i)$$

for all n and for almost all sample sequences $(x_1, \cdots, x_n)$, so

$$(2.5)\qquad P[\sup_{|\theta-\theta_0|\le\delta_0} \left|\frac{1}{n}\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2}\right| \le -I(\theta_0)+\eta_1+\frac{\epsilon}{n}\sum_{i=1}^{n} M(X_i) \text{ fasln}] = 1.$$

Another application of the SLLN implies that for any $\eta_2 > 0$,

$$P[\frac{1}{n}\sum_{i=1}^{n} M(X_i) \le EM+\eta_2 \text{ fasln}] = 1,$$

and this can be combined with (2.5) (using (1.2)) to yield

$$P[\sup_{|\theta-\theta_0|\le\delta_0} \left|\frac{1}{n}\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2}\right| \le -I(\theta_0) + \eta_1 + \epsilon(EM+\eta_2)\text{fasln}] = 1.$$

Since $\epsilon < I(\theta_0)/EM$, we can choose $\eta_1 > 0$ and $\eta_2 > 0$ sufficiently small so that $-I(\theta_0) + \eta_1 + \epsilon(EM+\eta_2) < 0$, thus proving the theorem.

Remarks: A version of Theorem 4 under slightly stronger assumptions was stated without proof by LeCam (1953, p. 308).

It is shown in Appendix II that under the assumptions C4, C5, C6, and (2.1), any weakly consistent root of the LEQ is asymptotically normal and asymptotically efficient. These assumptions are weaker than those usually made.

From Theorem 4 and Proposition 1, it follows that for almost all sample sequences $(x_1, \cdots)$ there is an integer $N(x_1, \cdots)$ such that for $n \geq N$, there is exactly one solution of the LEQ in I_{δ_0}. This solution is a relative maximum and converges to θ_0 as $n \to \infty$. If, furthermore, the MLE exists and is strongly consistent for θ_0 (see Theorems 9, 11, 12 of Section 3, or Wald (1949), LeCam (1953), or Perlman (1969) for conditions guaranteeing this) then the solution lying in I_{δ_0} must coincide with the MLE fasln.

The following conditions, similar to those commonly assumed in the literature, are stronger than C6 but may be easier to verify:

C6a. With $M(x)$ as in C6, there is a number $\alpha > 0$ such that for θ in I,

$$\sup_I \left[\left| \frac{\partial^2 \log p(x, \theta)}{\partial \theta^2} - \frac{\partial^2 \log p(x, \theta_0)}{\partial \theta^2} \right| - |\theta - \theta_0|^\alpha M(x) \right] \leq 0$$

for almost all x.

C6b. $(\partial^3/\partial\theta^3) \log p(x, \theta)$ exists for all θ in I and

$$\sup_I \left| \frac{\partial^3 \log p(x, \theta)}{\partial \theta^3} \right| \leq M(x)$$

for almost all x.

C6c. $(\partial^2/\partial\theta^2) \log p(x, \theta)$ is continuous at θ_0 uniformly in x.

3. Global limiting behavior of $S(x_1, \cdots)$.

In this section we study the behavior of the solutions of the LEQ over the entire range Ω. We need the following assumptions:

C8. For almost all x, $(\partial/\partial\theta) \log p(x, \theta)$ exists for all θ in Ω. (See the remark following C10).

C9. The integral

$$g_1(\theta) \equiv \int [\frac{\partial \log p(x, \theta)}{\partial\theta}] p(x, \theta_0) d\mu(x) = E[\frac{\partial \log p(X, \theta)}{\partial\theta}]$$

is finite at $\theta = \theta_0$. That is, the integral $\int [(\partial/\partial\theta) p(x, \theta_0)] d\mu(x)$ is assumed to exist.

C10. For each θ in Ω and for all $\varepsilon > 0$, there exists a nonnegative function $H_\theta(x)$ with $E_{\theta_0}[H_\theta(X)] < \infty$ and there exists $\delta(\varepsilon, \theta) > 0$ such that

$$\sup_{\{\theta' : |\theta' - \theta| \le \delta(\varepsilon,\theta)\}} \left| \frac{\partial \log p(x, \theta')}{\partial\theta} - \frac{\partial \log p(x, \theta)}{\partial\theta} \right| \le \varepsilon H_\theta(x)$$

for all x in Λ^c, where Λ is a set of probability 0 which does not depend on θ.

Remarks: Let $A_\theta = \{x: p(x, \theta) \ne 0, \infty\}$. If C8 is replaced by the weaker assumption C8': for almost all (P_{θ_0}) x, $\log p(x, \theta)$ exists, i.e., $p(x, \theta) \ne 0, \infty$, then this implies that

$$P_{\theta_0}[\bigcap_{\Omega} A_\theta] = 1.$$

In applications of the results of this paper, the true value θ_0 is unknown, so C8 (and the other conditions) must be assumed to hold for all θ_0 in Ω. Thus the set $\cap A_\theta$ has probability 1 under all possible distributions. Therefore, we may as well assume that for all x and all θ, $p(x, \theta) \ne 0, \infty$.

In Appendix III it is shown that if C8, C9, and C10 hold, then $g_1(\theta)$ exists for all θ in Ω. Furthermore, the Kullback-Leibler

(K-L) information number

$$I(\theta_0, \theta) = \int \log[\frac{p(x, \theta_0)}{p(x, \theta)}]p(x, \theta_0)d\mu(x)$$

is finite for all θ in Ω and

$$\frac{dI(\theta_0, \theta)}{d\theta} = - g_1(\theta).$$

Note that C10 implies that $(\partial/\partial\theta) \log p(x, \theta)$ and $g_1(\theta)$ are continuous functions of θ on Ω, for all x in Λ^c.

The following conditions are stronger than C10 but may be easier to verify:

C10a. For all θ in Ω there exist functions $H_\theta(x)$ as in C10 and an $\alpha > 0$ such that

$$\sup_{\theta' \in \Omega} [|\frac{\partial \log p(x, \theta')}{\partial\theta} - \frac{\partial \log p(x, \theta)}{\partial\theta}| - |\theta'-\theta|^\alpha H_\theta(x)] \le 0$$

for all x in Λ^c.

C10b. $(\partial^2/\partial\theta^2) \log p(x, \theta)$ exists for all θ in Ω and

$$\sup_\Omega |\frac{\partial^2 \log p(x, \theta)}{\partial\theta^2}| \le H(x)$$

for all x in Λ^c, where $H(x)$ is a non-negative function with $E_{\theta_0}[H(X)] < \infty$.

C10c. $(\partial/\partial\theta) \log p(x, \theta)$ is continuous at each θ, uniformly in x (for x in Λ^c).

The following lemma is a form of Ascoli's Theorem.

Lemma 5. Let $\{f_n(\theta)\}$ be a sequence of continuous real-valued functions defined on a compact subset K of the real line, such that for each θ in K, $f_n(\theta) \to f(\theta)$ as $n\to\infty$. Then $f_n(\theta) \to f(\theta)$ uniformly on K if and only if the sequence $\{f_n(\theta)\}$ is equicontinuous on K, i.e., for each θ in K and each $\epsilon > 0$ there exists $\delta(\epsilon, \theta) > 0$ such that $|\theta'-\theta| \leq \delta(\varepsilon, \theta)$ implies $|f_n(\theta')-f_n(\theta)| \leq \varepsilon$ for all n.

Proof. The "if" assertion is proved in Royden (1968, p. 178, Lemma 32). The converse follows easily from the inequality

$$|f_n(\theta')-f_n(\theta)| \leq |f_n(\theta')-f(\theta')| + |f(\theta')-f(\theta)| + |f(\theta)-f_n(\theta)|.$$

This lemma is used to prove the following theorem, which is similar to a result of Rubin (1956).

THEOREM 6. If C8, C9, and C10 are true and if K is a compact subset of Ω, then

$$P[\sup_K |\frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial \theta} - g_1(\theta)| \to 0] = 1.$$

Proof. For each sample sequence $(x_1, \cdots)$ define the sequence of functions $\{f_n(\theta)\} = \{f_n(\theta, x_1, \cdots, x_n)\}$ by

$$f_n(\theta) = \frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log p(x_i, \theta)}{\partial \theta} .$$

We first show that

$$(3.1) \qquad P[\{f_n(\theta)\} \text{ is equicontinuous on } \Omega] = 1.$$

For any $\theta \in \Omega$ the SLLN implies that

$$P[\frac{1}{n} \sum_{i=1}^{n} H_\theta(X_i) \leq EH_\theta+1 \text{ fasln}] = 1,$$

Given any $\varepsilon' > 0$, let $\varepsilon = \varepsilon'(EH_\theta+1)^{-1}$. By C10, there is a $\delta(\varepsilon, \theta)$ such that $|\theta'-\theta| \leq \delta(\varepsilon, \theta)$ implies that for all n

$$|f_n(\theta')-f_n(\theta)| \leq \frac{\varepsilon}{n} \sum_{i=1}^{n} H_\theta(x_i).$$

With probability 1 this last expression is $\leq \varepsilon'$ fasln. Since, for any finite N the family $\{f_n(\theta)\}_{n=1}^{N}$ is equicontinuous, this proves (3.1).

Next let $D = \{\theta_k\}$ be any countable set dense in K. For each k the SLLN implies that

$$P[f_n(\theta_k) \to g_1(\theta_k)] = 1,$$

so

$$P[f_n(\theta_k) \to g_1(\theta_k) \text{ for all } k] = 1.$$

Then by (3.1), it follows from a well-known result (see Royden, p. 178, Lemma 31) that

$$P[f_n(\theta) \to g_1(\theta) \text{ for all } \theta \text{ in } \Omega] = 1. \tag{3.2}$$

Hence by Lemma 5 and (3.1)

$$P[f_n(\theta) \to g_1(\theta) \text{ uniformly on } K] = 1,$$

which proves the theorem.

From the theorem it follows that if G is a compact set such that

$$\inf_G |g_1(\theta)| > 0, \tag{3.3}$$

then

$$P[\inf_G \left|\frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial\theta}\right| > 0 \text{ fasln}] = 1. \tag{3.4}$$

so that with probability 1 there exist no solutions of the LEQ in G fasln. Hence Theorems 4 and 6 may be combined to yield

<u>THEOREM</u> <u>7</u>. Suppose conditions C1, 4, 5, 7, 8, 9, and 10 are true. If G is a compact subset of Ω satisfying (3.3) and if $K = G \cup I_\delta$ where $0 < \delta \le \delta_0$, then

$$P[\exists \text{ exactly one solution of the LEQ in } K \text{ fasln}] = 1. \tag{3.5}$$

This result is most useful when θ_0 is an interior point of the compact interval K, where now δ is chosen so that $I_\delta \subset K$ and

$$G = \text{closure } (K - I_\delta\ ). \tag{3.6}$$

Now, it is well-known that under the assumption C1, the K-L number $I(\theta_0, \theta)$ is strictly positive if $\theta \neq \theta_0$, so $I(\theta_0, \theta)$ achieves its minimum value (zero) at $\theta = \theta_0$. By the remark following C10 this implies $g_1(\theta_0) = 0$. If we assume that $g_1(\theta)$ has no zeroes in K other than that at θ_0, then (3.3) holds since $g_1(\theta)$ is continuous and G is compact. Now consider the

<u>Definition</u>: Let $J \subset \Omega$ be an interval containing θ_0. The family $\{p(x, \theta): \theta \in J\}$ is <u>Kullback-Leibler well-ordered</u> (briefly, <u>K-L ordered</u>) <u>at θ_0</u> if $0 < I(\theta_0, \theta_1) < I(\theta_0, \theta_2)$ whenever $\theta_0 < \theta_1 < \theta_2$ or $\theta_2 < \theta_1 < \theta_0$ (θ_1, θ_2 in J). The family is <u>strongly Kullback-Leibler well-ordered</u> (<u>strongly K-L ordered</u>) <u>at θ_0</u> if, for θ in J,

$$\frac{dI(\theta_0, \theta)}{d\theta} \begin{cases} < 0 & \text{if } \theta < \theta_0 \\ > 0 & \text{if } \theta > \theta_0 \end{cases} .$$

(At any boundary point of J take the one-sided derivative.) The family is (<u>strongly</u>) <u>K-L ordered</u> if it is (strongly) K-L ordered at each θ_0 in Ω.

Therefore, if we assume that $\{p(x, \theta): \theta \in K\}$ is strongly K-L ordered at θ_0, then for θ in K, $g_1(\theta) > 0 (< 0)$ if $\theta < \theta_0 (\theta > \theta_0)$. Hence $g_1(\theta)$ has no zeroes in K other than at θ_0, so (3.3) holds for the set G in (3.3), and so (3.5) holds for K. If $\{p(x, \theta): \theta \in \Omega\}$ is strongly K-L ordered at θ_0, then (3.5) holds for all compact intervals $K \subset \Omega$ such that θ_0 is an interior point of K. In this case, with probability 1 all solutions of the LEQ except one must approach the boundary $\{a\} \cup \{b\}$ of Ω as $n \to \infty$. We summarize these results in

<u>THEOREM 8</u>. If conditions C1, 4, 5, 7, 8, 9, and 10 are true and if $\{p(x, \theta): \theta \in \Omega\}$ is strongly K-L ordered at θ_0, then (3.5) holds for any compact interval $K \subset \Omega$ with $\theta_0 \in$ interior (K). Then for almost all sample sequences $(x_1, \cdots)$ and any neighborhoods U, V, and I_δ of a, b, and θ_0, respectively $(\delta \leq \delta_0)$, there exists $N(x_1, \cdots)$ (depending on U, V, δ) such that for all $n \geq N(x_1, \cdots)$,

$$S(x_1, \cdots, x_n) \subset U \cup V \cup I_\delta$$

and exactly one member of $S(x_1, \cdots, x_n)$ lies in I_δ. This exceptional solution is a relative maximum and converges to θ_0 as $n \to \infty$. It coincides with the MLE fasln if the latter exists and is strongly consistent for θ_0. (By a neighborhood of $-\infty$ or $+\infty$ is meant an interval of the form $(-\infty, y)$ or $(y, +\infty)$, respectively.)

If Ω itself is a compact interval and if the assumptions of Theorem 8 hold, then (3.5) holds with K replaced by Ω. In this case, with probability 1 there exists a unique solution of the LEQ in Ω fasln, so this solution must give the <u>unique absolute maximum</u> of the LF. Thus we have

<u>THEOREM 9</u>. Suppose Ω is a compact interval. If C1, 4, 5, 7, 8, 9, and 10 are true and if $\{p(x, \theta): \theta \in \Omega\}$ is strongly K-L ordered at θ_0, then (3.5) holds with $K = \Omega$. Thus for almost all sample sequences $(x_1, \cdots)$ there exists $N(x_1, \cdots)$ such that for all $n \geq N(x_1, \cdots)$,

$S(x_1, \cdots, x_n)$ contains exactly one member. This solution of the LEQ is the unique absolute maximum of the LF and converges to θ_0 as $n \to \infty$.

Remark: If Ω is compact, C8 and C10 imply that the MLE exists for all n and $(x_1, \cdots, x_n)$. We have proved that under the assumptions of Theorem 9 the MLE is strongly consistent for θ_0. The assumptions of Wald (1949) are weaker than ours and do not imply the existence of the MLE. See Theorems 11 and 12, and also Perlman (1969) for more general results.

To obtain the analog of Theorem 9 in the case where Ω is not compact, we must extend Lemma 5. We assume for the remainder of Section 3 that Ω is an open interval (a, b) with $-\infty \leq a < b \leq +\infty$. (The following argument can easily be combined with the preceding work to treat the case where Ω is half-open.)

Let $\{f_n(\theta)\}$ be a sequence of continuous real-valued functions on Ω such that $f_n(\theta) \to f(\theta)$ for each θ in Ω. Let $h_n(\theta) = f_n(\theta) - f(\theta)$,

$$a_n = \liminf_{\theta \to a} h_n(\theta), \quad A_n = \limsup_{\theta \to a} h_n(\theta),$$

$$b_n = \liminf_{\theta \to b} h_n(\theta), \quad B_n = \limsup_{\theta \to b} h_n(\theta),$$

and denote the closed intervals $[a_n, A_n]$ and $[b_n, B_n]$ by J_n and Q_n respectively. For any real number x and any set $A \subset (-\infty, \infty)$, let $|x - A| = \inf_{\{y \in A\}} |x - y|$.

Lemma 10. With $\{f_n(\theta)\}$ as given above, $f_n(\theta) \to f(\theta)$ uniformly on Ω if and only if

(a) $\{f_n(\theta)\}$ is equicontinuous on Ω,

(b) $A_n - a_n \to 0$ and $B_n - b_n \to 0$ as $n \to \infty$, and

(c) for each $\epsilon > 0$ there exist open intervals $U(\epsilon)$ and $V(\epsilon)$, containing a and b respectively, such that θ in $U(\epsilon) \cap \Omega$ implies $|h_n(\theta) - J_n| \le \epsilon$ for all n, and θ in $V(\epsilon) \cap \Omega$ implies $|h_n(\theta) - Q_n| \le \epsilon$ for all n.

Proof. First we show that (b) and (c) imply

(d) $A_n \to 0$, $a_n \to 0$, $B_n \to 0$, and $b_n \to 0$ as $n \to \infty$.

Given $\epsilon > 0$, choose any $\bar{\theta}$ in $U(\epsilon)$, so that $|h_n(\bar{\theta}) - J_n| \le \epsilon$ for all n. If we choose n_0 so that $n \ge n_0$ implies $A_n - a_n \le \epsilon$, then $n \ge n_0$ implies $|h_n(\bar{\theta}) - A_n| \le 2\epsilon$. However, $|A_n| \le |h_n(\bar{\theta})| + |h_n(\bar{\theta}) - A_n|$ and $h_n(\bar{\theta}) \to 0$ as $n \to \infty$, so $A_n \to 0$. The other limits are obtained similarly.

Now assume that (a), (b) and (c) hold. Given $\epsilon > 0$ the set $K = \Omega - (U(\epsilon) \cup V(\epsilon))$ is a compact subinterval of Ω, so by Lemma 5 there is an n_1 such that for all $n \ge n_1$,

$$\sup_K |h_n(\theta)| \le \epsilon.$$

By (d) there is an integer n_2 such that for all $n \ge n_2$, $|A_n| \le \epsilon$ and $|a_n| \le \epsilon$. Then for any θ in $U(\epsilon) \cap \Omega$, $n \ge n_2$ implies that

$$\min[h_n(\theta) - A_n,\ a_n - h_n(\theta),\ |h_n(\theta)|] \le \epsilon,$$

so $|h_n(\theta)| \le 2\epsilon$. Hence $n \ge n_2$ implies

$$\sup_{U(\epsilon)\cap\Omega} |h_n(\theta)| \le 2\epsilon$$

and similarly we can find n_3 such that $n \ge n_3$ implies

$$\sup_{V(\epsilon)\cap\Omega} |h_n(\theta)| \le 2\epsilon.$$

Thus $n \ge \max(n_1, n_2, n_3)$ implies that $|h_n(\theta)| \le 2\epsilon$ for all θ in Ω, so $h_n(\theta) \to 0$ uniformly on Ω.

Conversely, suppose $f_n(\theta) \to f(\theta)$ uniformly on Ω. Lemma 5 implies (a), since Ω can be expressed as a union of compact subintervals. Next, given $\varepsilon > 0$, choose n_4 such that $\sup|h_n(\theta)| \le \frac{\varepsilon}{3}$ for all $n \ge n_4$. Hence $n \ge n_4$ implies that $|A_n|$, $|a_n|$, $|B_n|$, and $|b_n|$ are all $\le \frac{\varepsilon}{3}$, proving (b). Finally, for $n \ge n_4$ and for all θ in Ω

$$|h_n(\theta) - J_n| \le |h_n(\theta)| + |A_n| + |a_n| \le \varepsilon.$$

We can now find an open interval $U(\varepsilon)$ containing a such that $\theta \in U(\varepsilon) \cap \Omega$ implies $|h_n(\theta) - J_n| \le \varepsilon$ for $n = 1, \cdots, n_4$, so (c) is true. This completes the proof of the lemma.

Remark: Suppose that

(b') for all n (or fasln) $\lim\limits_{\theta \to a} h_n(\theta) \equiv h_n(a)$ and $\lim\limits_{\theta \to b} h_n(\theta) \equiv h_n(b)$ both exist.

Then condition (b) is satisfied trivially and (c) can be restated in the form

(c') $\lim\limits_{\theta \to a} h_n(\theta) = h_n(a)$ uniformly in n and

$\lim\limits_{\theta \to b} h_n(\theta) = h_n(b)$ uniformly in n.

To apply these results, for each sample sequence $(x_1, \cdots)$ define

$$f_n(\theta) = f_n(\theta, x_1, \cdots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log p(x_i, \theta)}{\partial \theta} = \frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial \theta},$$

let $f(\theta) = g_1(\theta)$, and let $h_n(\theta)$, A_n, a_n, B_n, b_n, J_n, and Q_n be as defined before Lemma 10. Notice that these quantities depend on $(x_1, \cdots)$. Now assume that C8, C9, and C10 hold and that either of the following two conditions are true:

(3.7) $$P[\text{(b) and (c) are satisfied}] = 1$$

(3.8) $$P[\text{(b') and (c') are satisfied}] = 1.$$

That (a) holds with probability 1 follows from (3.1). Then the proof of Theorem 6 may be carried over virtually without change (but replacing Lemma 5 by Lemma 10) to show that

(3.9) $$P[\sup_{\Omega}\left|\frac{1}{n}\frac{\partial \log L_n(\theta)}{\partial\theta} - g_1(\theta)\right| \to 0] = 1.$$

We now show that (3.8) is implied by

C11. Let $m(x, \theta) = [\partial/\partial\theta \log p(x, \theta)] - g_1(\theta)$. For almost all x, $\lim_{\theta\to a} m(x, \theta) \equiv m(x, a)$ and $\lim_{\theta\to b} m(x, b) \equiv m(x, b)$ exist. Furthermore, there exists nonnegative functions $H_a(x)$ and $H_b(x)$, with $E_{\theta_0}[H_a(X)] < \infty$ and $E_{\theta_0}[H_b(X)] < \infty$, and for each $\epsilon > 0$ there exist open intervals $Y(\epsilon)$, $Z(\epsilon)$ containing a and b respectively, such that for almost all x

$$\sup_{Y(\epsilon)\cap\Omega} |m(x, \theta) - m(x, a)| \le \epsilon H_a(x),$$

$$\sup_{Z(\epsilon)\cap\Omega} |m(x, \theta) - m(x, b)| \le \epsilon H_b(x).$$

If C11 is true, then (b') is satisfied with probability 1 since $h_n(\theta) = n^{-1}\sum_{i=1}^{n} m(x_i, \theta)$. The SLLN implies that for almost all sample sequences $(x_1, \cdots)$, $n^{-1}\sum_{i=1}^{n} H_a(x_i) \le EH_a + 1$ if $n \ge N(x_1, \cdots)$. Given $\epsilon' > 0$ let $\epsilon = \epsilon'(EH_a+1)^{-1}$. Then $\theta \in Y(\epsilon) \cap \Omega$ implies

$$|h_n(\theta) - h_n(a)| \le \frac{1}{n}\sum_{i=1}^{n} |m(x_i, \theta) - m(x_i, a)|$$

$$\le \frac{\epsilon}{n}\sum_{i=1}^{n} H_a(x_i),$$

and the last term is $\le \epsilon$ if $n \ge N(x_1, \cdots)$. Thus (c') is satisfied with probability 1.

Note that a condition similar to but weaker than C11, involving inferior and superior limits of $m(x, \theta)$, will suffice to guarantee (3.7). Also, the following condition is stronger than C11 but may be easier to verify:

C11a. The limits $\lim_{\theta \to a} m(x, \theta)$ and $\lim_{\theta \to b} m(x, \theta)$ exist and are uniform in x.

If we now assume that $\{p(x, \theta) : \theta \in \Omega\}$ is strongly K-L ordered at θ_0 and that for some constant $A > 0$

$$(3.10) \qquad \liminf_{\theta \to a, b} \left|\frac{dI(\theta_0, \theta)}{d\theta}\right| \geq A,$$

then for any $\delta > 0$

$$(3.11) \qquad \inf_{\Omega - I_\delta} |g_1(\theta)| > 0.$$

Then by (3.9)

$$(3.12) \qquad P[\exists \text{ no solutions of the LEQ in } \Omega - I_\delta \text{ fasln}] = 1.$$

Combining this with Theorem 4 we have

THEOREM 11. Let Ω be an open interval (a, b), possibly unbounded. Suppose that conditions C1, 4, 5, 7, 8, 9, 10, and 11 are true, that $\{p(x, \theta): \theta \in \Omega\}$ is strongly K-L ordered at θ_0, and that (3.10) is satisfied. Then (3.5) holds with $K = \Omega$. Therefore, for almost all sample sequences $(x_1, \cdots)$ there exists $N(x_1, \cdots)$ such that for all $n \geq N$, $S(x_1, \cdots, x_n)$ contains exactly one member. This unique solution of the LEQ is the unique absolute maximum of the LF and converges to θ_0 as $n \to \infty$. Hence, with probability 1 the MLE exists fasln and is strongly consistent.

If C11 (or a similar condition) fails then uniform convergence of $f_n(\theta)$ over the entire range Ω (i.e., the conclusion of Theorem 6 with $K = \Omega$) need not occur. However, the conclusion of Theorem 11 remains true if C11 is replaced by

C12. There exist functions $W_a(x)$ and $W_b(x)$, with $E_{\theta_0}[W_a(X)] > 0$ and $E_{\theta_0}[W_b(X)] < 0$, and there exist open intervals Y, Z containing a, b respectively, such that for almost all x

$$\inf_{Y \cap \Omega} \frac{\partial \log p(x, \theta)}{\partial \theta} \geq W_a(x)$$

$$\sup_{Z \cap \Omega} \frac{\partial \log p(x, \theta)}{\partial \theta} \leq W_b(x).$$

Theorem 12. Let Ω be an open interval (a, b), possibly unbounded. Suppose that conditions C1, 4, 5, 7, 8, 9, 10, and 12 are true, and that $\{p(x, \theta) : \theta \in \Omega\}$ is strongly K-L ordered at θ_0. Then the conclusions of Theorem 11 are valid.

Proof. C12 implies that (3.10) and (3.11) hold, but we do not use this fact. The assumptions made here are sufficient to imply that (3.5) holds for the compact set $K = \Omega - (Y \cap Z)$. By a now-familiar application of the SLLN, C12 implies that

$$P[\inf_{Y \cap \Omega} \frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial \theta} > 0 \text{ fasln}] = 1$$

$$P[\sup_{Z \cap \Omega} \frac{1}{n} \frac{\partial \log L_n(\theta)}{\partial \theta} < 0 \text{ fasln}] = 1,$$

so

$$P[\exists \text{ no solutions of the LEQ in } \Omega - K \text{ fasln}] = 1.$$

This, together with (3.5), proves the result.

Remark: Suppose that all conditions of Theorem 11 are satisfied except C11. If there is a bounded continuous function $k(\theta) > 0$ on Ω such that C11 (or the weaker condition involving superior and inferior limits) is satisfied when $m(x, \theta)$ is replaced by $m^*(x, \theta) = k(\theta)m(x, \theta)$, then the preceding argument carries over if we substitute $f_n^*(\theta) = k(\theta)f_n(\theta)$ for $f_n(\theta)$, $g_1^*(\theta) = k(\theta)g_1(\theta)$ for $g_1(\theta)$, and $h_n^*(\theta) = k(\theta)h_n(\theta)$ for $h_n(\theta)$, so (3.9) is replaced by

$$(3.13) \qquad P\left[\sup_{\Omega}\left|\frac{1}{n}\, k(\theta)\, \frac{\partial \log L_n(\theta)}{\partial\theta} - g_1^*(\theta)\right| \to 0\right] = 1.$$

If assumption (3.10) is replaced by

$$(3.14) \qquad \liminf_{\theta \to a,b} |g_1^*(\theta)| = \liminf_{\theta \to a,b} k(\theta)|g_1(\theta)| \geq A > 0,$$

then (3.11) holds with $g_1(\theta)$ replaced by $g_1^*(\theta)$. This, together with (3.13) and the fact that $k(\theta) > 0$, implies (3.12) so the conclusion of Theorem 11 remains true in this case. This situation is illustrated in Section 5 in the problem of estimating the correlation in a bivariate normal population with known variances.

Similarly, if all conditions of Theorem 12 except C12 are satisfied, but if C12 is satisfied when $(\partial/\partial\theta) \log p(x,\theta)$ is replaced by $k(\theta)(\partial/\partial\theta) \log p(x, \theta)$, then the conclusion of Theorem 12 remains true.

4. Some examples of Kullback-Leibler well-ordered families.

Let $\{p(x, \theta): \theta \in \Omega\}$ be a family of densities satisfying conditions C1, C8, C9, and C10, for all θ_0 in Ω. (These assumptions are weakened in Propositions 14 and 16.) We present several sufficient conditions for this family to be K-L ordered or strongly K-L ordered.

First, if $(\partial/\partial\theta) \log p(x, \theta)$ is a decreasing function of θ for almost all x, then $g_1(\theta)$ is also decreasing in θ. Since $g_1(\theta) = 0$ this implies that $g_1(\theta) \geq (\leq)\ 0$ if $\theta \leq (\geq)\ \theta_0$. If, furthermore, C4, C5, and C7 hold for all θ_0 in Ω then $g_1'(\theta_0) = - I(\theta_0) < 0$ (see Appendix III), so the preceding inequalities are strict. Thus the family is strongly K-L ordered, and in addition (3.10) is satisfied for all θ_0 in Ω, since $g_1(\theta)$ is decreasing. If C4, C5, and/or C7 are not satisfied, but if $(\partial/\partial\theta) \log p(x, \theta)$ is strictly decreasing in θ for all x in a set of positive measure, then $g_1(\theta)$ is strictly decreasing and the conclusions in the preceding sentence remain true.

Before presenting more interesting sufficient conditions, recall (Hardy, Littlewood, and Polya (1952) p. 168) that two real-valued functions $\phi(x)$, $\psi(x)$ defined on a common domain are <u>similarly ordered</u> (S.O.) if

$$[\phi(x) - \phi(y)]\,[\psi(x) - \psi(y)] \geq 0$$

for all $x \neq y$. ϕ and ψ are <u>strictly similarly ordered</u> if strict inequality holds for all $x \neq y$. ϕ and ψ are <u>oppositely ordered</u> (O.O.) if

$$[\phi(x) - \phi(y)]\,[\psi(x) - \psi(y)] \leq 0$$

for all $x \neq y$ and are <u>strictly oppositely ordered</u> if strict inequality holds for all $x \neq y$. If ϕ and ψ are S.O. and if Z is any random variable taking values in the common domain of ϕ and ψ, then

$$E[\phi(Z)\psi(Z)] \geq E[\phi(Z)]E[\psi(Z)], \tag{4.1}$$

and the reverse inequality holds if ϕ and ψ are O.O. Strict inequality holds if ϕ and ψ are strictly S.O. or strictly O.O. and Z is a non-degenerate random variable.

Consider the following versions of the monotone likelihood ratio property (also, we can replace "increasing" by "decreasing" throughout):

(i)' For all θ, $(\partial/\partial\theta) \log p(x, \theta)$ is an increasing function of a real valued statistic $T(x)$.

(ii)' For all $\theta_1 < \theta_2$, $p(x, \theta_2)/p(x, \theta_1)$ is an increasing function of $T(x)$.

(iii)' For all $\theta_1 < \theta_2$, $p(x, \theta_2)/p(x, \theta_1)$ is a strictly increasing function of $T(x)$.

(iv)' For all θ, $(\partial/\partial\theta) \log p(x, \theta)$ is a strictly increasing function of $T(x)$.

It follows from Theorems 2.1, 2.3, and 2.4 of Karlin (1968, Chapter 2) that (iv)' => (iii)' => (ii)' => (i)' (assuming differentiability), and furthermore all of the reverse implications are false in general. In the terminology used by Karlin (p. 49), (iv)' states that $p(x, \theta)$ is $ETP_2(\theta)$. We now generalize these conditions as follows:

(i) For all $\theta_1 < \theta_2$, the functions $(\partial/\partial\theta) \log p(x, \theta_1)$ and $p(x, \theta_2)/p(x, \theta_1)$ are S.O.

(ii) For all $\theta_1 < \theta_2 < \theta_3$, the functions $p(x, \theta_2)/p(x, \theta_1)$ and $p(x, \theta_3)/p(x, \theta_2)$ are S.O.

(iii) There is a function $T(x)$ such that for all $\theta_1 < \theta_2 < \theta_3$, $p(x, \theta_2)/p(x, \theta_1)$ and $p(x, \theta_3)/p(x, \theta_2)$ are strictly S.O. functions of $T(x)$.

(iv) For all $\theta_1 < \theta_2$, $(\partial/\partial\theta) \log p(x, \theta_1)$ and $p(x, \theta_2)/p(x, \theta_1)$ are strictly S.O. functions of $T(x)$.

Clearly, (i)' => (i), (ii)' => (ii), (iii)' => (iii), and (iv)' => (iv). Furthermore, an extension of the results given by Karlin shows that (iv) => (iii) => (ii) => (i).

<u>Proposition 13</u>. Suppose that the family $\{p(x, \theta): \theta \in \Omega\}$ satisfies (iv) and conditions C1, C8, C9, and C10, for all θ_0 in Ω. Then this family is strongly K-L ordered.

<u>Proof</u>. For any $\theta_1 > \theta_0$ we must show that $g_1(\theta_1) < 0$. Now

$$(4.2)\quad g_1(\theta_1) = \int[\frac{\partial \log p(x,\ \theta_1)}{\partial\theta}]\ p(x,\ \theta_0)d\mu(x)$$

$$= \int[\frac{\partial \log p(x,\ \theta_1)}{\partial\theta}]\ \frac{p(x,\ \theta_0)}{p(x,\ \theta_1)}\ p(x,\ \theta_1)d\mu(x)$$

$$< \{\int[\frac{\partial \log p(x,\ \theta_1)}{\partial\theta}]\ p(x,\ \theta_1)d\mu(x)\}\{\int\frac{p(x,\ \theta_0)}{p(x,\ \theta_1)}\ p(x,\ \theta_1)d\mu(x)\}$$

$$= E_{\theta_1}[\frac{\partial \log p(X,\ \theta_1)}{\partial\theta}] = 0.$$

The inequality follows from the remark after (5.1) and the fact that $(\partial/\partial\theta) \log p(x, \theta_1)$ and $p(x, \theta_0)/p(x, \theta_1)$ are strictly O.O. functions of $T(x)$. $T(X)$ must be a non-degenerate random variable: if not, then the function $h(x) \equiv p(x, \theta_0)/p(x, \theta_1)$, being a function of $T(x)$, must be constant for almost all x, say $h(x) \equiv c$. Since $p(x, \theta_0)$ and $p(x, \theta_1)$ are both densities, $c = 1$, but this contradicts C1. Lastly, the final equality in (4.2) is true because C8, C9, and C10 are assumed to hold with θ_0 replaced by θ_1. The proof that $g_1(\theta_1) > 0$ if $\theta_1 < \theta_0$ is identical.

Remark: A sufficient condition for (iv) (in fact, for (iv)') to hold with $T(x) = x$ is that $(\partial^2/\partial x\partial\theta) \log p(x, \theta) > 0$ for all θ and almost all x, assuming that this mixed derivative exists. Also, (iv) is satisfied if this inequality is reversed.

The next proposition is not needed in the present paper, but will be used in subsequent papers, e.g., Perlman (1969).

Proposition 14. Suppose that the family $\{p(x, \theta) : \theta \in \Omega\}$ satisfies (ii), C1, and

(a) $I(\theta_0, \theta)$ is finite for all θ_0 and θ in Ω.

Then this family is K-L ordered. Thus, a monotone likelihood ratio family is K-L ordered.

Proof. It suffices to show that for all $\theta_0 < \theta_1 < \theta_2$,

$$J \equiv \int \log\left[\frac{p(x, \theta_1)}{p(x, \theta_2)}\right] p(x, \theta_0)\, d\mu(x) > 0.$$

(The proof for the case $\theta_2 < \theta_1 < \theta_0$ is similar). Now

$$J = \int \log\left[\frac{p(x, \theta_1)}{p(x, \theta_2)}\right] \frac{p(x, \theta_0)}{p(x, \theta_1)}\; p(x, \theta_1) d\mu(x)$$

$$\geq \{\int \log\left[\frac{p(x, \theta_1)}{p(x, \theta_2)}\right] p(x, \theta_1) d\mu(x)\}\{\int \frac{p(x, \theta_0)}{p(x, \theta_1)}\, p(x, \theta_1) d\mu(x)\}$$

$$= I(\theta_1, \theta_2).$$

The inequality follows from (5.1) and (ii). However, the K-L information number $I(\theta_1, \theta_2)$ is strictly positive, by C1 and the information inequality, so $J > 0$, completing the proof.

We now specialize to the case where $\{p(x, \theta): \theta \in \Omega\}$ is a location parameter family, i.e., $p(x, \theta) = f(x-\theta)$. Here x is a real variable, f is a density (with respect to Lebesgue measure) on the real line, and $\Omega = (-\infty, \infty)$. In this case condition C1 is always satisfied, condition C8 is equivalent to the requirement that for all x, $f(x) > 0$ and $f'(x)$ exists, and C9 is equivalent to the existence of the integral $\int_{-\infty}^{\infty} f'(x)dx$. Note that in these forms the conditions do not depend on θ_0, so if C8 and C9 hold for one θ_0 they hold for all θ_0. Similarly the other C-conditions can be stated in simplified forms, e.g., C10c now requires that $f'(x)/f(x)$ be uniformly continuous on $(-\infty, \infty)$.

Propositions 13 and 14 give sufficient conditions under which the family $\{f(x-\theta): -\infty < \theta < \infty\}$ is strongly K-L ordered or K-L ordered.

In particular, Proposition 13 implies that if this family satisfies C8, C9, C10 (for all θ_0), and (iv)' (with $T(x) = (x)$) then it is strongly K-L ordered. We can, in fact, conclude that here (3.10) holds as well, for (iv)' implies that $(\partial/\partial\theta) \log f(x-\theta)$ is a strictly decreasing function of θ, so the remarks contained in the second paragraph of this section can be applied. Similarly, if we make the additional assumption that C4, C5, and C7 hold (for all θ_0), then (iv)' can be replaced by the weaker condition (i)' and we again conclude that the family is strongly K-L ordered and satisfies (3.10).

We now show that Propositions 13 and 14 remain true if the "generalized monotone likelihood ratio" conditions appearing there are replaced by the assumption that $f(x)$ is a symmetric unimodal density (with mode at 0). That is, assume that $f(x) = f(-x)$ and that $f(x)$ is increasing on $(-\infty, 0]$ and decreasing on $[0, \infty)$. Strict monotonicity is not assumed.

Proposition 15. Suppose that the family $\{f(x-\theta): -\infty < \theta < \infty\}$ satisfies C8, C9, and C10 (for all θ_0) and that $f(x)$ is symmetric and unimodal. Then this family is strongly K-L ordered.

Proof. Without loss of generality take $\theta_0 = 0$. Letting $h(x) = f'(x)/f(x)$, the function $(d/d\theta)I(0, \theta) = -g_1(\theta)$ is given by

$$\int_{-\infty}^{\infty} h(x-\theta)f(x)dx = \int_{-\infty}^{\infty} h(x)f(x+\theta)dx.$$

Since $f(x)$ is an even (symmetric) function, $f'(x)$ and $h(x)$ are odd functions so

$$-g_1(\theta) = \int_0^{\infty} h(x)[f(x+\theta) - f(x-\theta)]dx.$$

Suppose that $\theta > 0$. Then for all $x > 0$, $f(x+\theta) - f(x-\theta) \le 0$ and $f'(x) \le 0$, by symmetry and unimodality, so $g_1(\theta) \le 0$. To see that this inequality is actually strict, consider the set $B_\theta = \{x: x \ge \theta, f'(x) < 0\}$.

Now, $f(x) > 0$ and $f'(x) \leq 0$ for all positive x, $f'(x)$ is continuous (by C10), and $f(x) \to 0$ as $x \to \infty$, so B_θ must have positive measure (otherwise $f(x)$ would be constant for $x \geq \theta$). Furthermore, by unimodality and symmetry, x in B_θ implies that $f(x+\theta) - f(x-\theta) < 0$, so $-g_1(\theta) \geq \int_{B_\theta} h(x)[f(x+\theta) - f(x-\theta)]dx > 0$.

Similarly, if $\theta < 0$ then $g_1(\theta) > 0$, completing the proof.

<u>Proposition 16</u>. Suppose that the family $\{f(x-\theta): -\infty < \theta < \infty\}$ satisfies (a) of Proposition 14 and that $f(x)$ is symmetric and unimodal. Then the family is K-L ordered.

<u>Proof</u>. Again take $\theta_0 = 0$ and choose $\theta_2 > \theta_1 > 0$. We must show that

$$J \equiv \int_{-\infty}^{\infty} \log[\frac{f(x-\theta_1)}{f(x-\theta_2)}]f(x)dx > 0.$$

Defining $\nu = \frac{1}{2}(\theta_1+\theta_2)$, $\delta = \frac{1}{2}(\theta_2-\theta_1)$, and $t(y) = \log[f(y+\delta)/f(y-\delta)]$, we find that

$$J = \int_{-\infty}^{\infty} t(x-\nu)f(x)dx = \int_{-\infty}^{\infty} t(x)f(x+\nu)dx.$$

Since $f(x)$ is symmetric, $t(y) = -t(-y)$ so

$$J = \int_0^{\infty} t(x)[f(x+\nu) - f(x-\nu)]dx.$$

Furthermore, since $\delta > 0$ and $\nu > 0$, the unimodality and symmetry of $f(x)$ implies that for all $x > 0$, $t(x) \leq 0$ and $f(x+\nu) - f(x-\nu) \leq 0$, so $J \geq 0$. To see that $J > 0$, consider the set $D_\nu = \{x: x \geq \nu, t(x) < 0\}$. Since $f(x)$ is positive and decreases to zero as $x \to \infty$, D_ν must have positive measure (otherwise $f(x)$ would be a constant). Also, since $0 < \delta < \nu$, x in D_ν implies that $f(x+\nu) - f(x-\nu) < 0$. Then

$$J \geq \int_{D_\nu} t(x)[f(x+\nu) - f(x-\nu)]dx > 0.$$

The proof for the case $\theta_2 < \theta_1 < 0$ is similar.

5. Some special densities.

If $\{p(x, \theta): \theta \in \Omega\}$ is a generalized exponential family, i.e., $p(x, \theta) = \beta(\theta) \exp [\theta T(x)]$, then all conditions of Theorem 11 are satisfied. It is easy to see, however, that for all n and $(x_1, \cdots, x_n)$ the LEQ has exactly one solution, a relative maximum, so the conclusions of Theorem 11 are valid *a fortiori*. Two non-trivial examples which nicely illustrate the results of the preceding sections are now presented.

Example 1: The Cauchy distribution with unknown location parameter.

For x real let $f(x) = \pi^{-1}(1+x^2)^{-1}$, $p(x, \theta) = f(x-\theta)$, μ = Lebesgue measure, and $\Omega = (-\infty, \infty)$. Recalling the remarks of Section 4 concerning location parameter families, it is easy to see that all conditions of Theorem 8 are satisfied, for all θ_0 in Ω. (Here, conditions C6b and C10b are easy to verify and imply C7 and C10 respectively. Proposition 15 implies that the family is strongly K-L ordered.) Thus, for any compact interval K of which θ_0 is an interior point, with probability 1 there is exactly one solution of the LEQ (a relative maximum) in K fasln, this solution converges to θ_0, and it coincides with the MLE fasln *if* the MLE is strongly consistent.

Next, note that $g_1(\theta) \to 0$ as $\theta \to \pm\infty$ (by the Bounded Convergence Theorem) so (3.10) fails and Theorem 11 does not apply. Of course, this does not guarantee the existence of multiple solutions of the LEQ; however, direct examination of the LEQ (see Barnett) shows that there may be as many as $2n$ solutions, of which n may give relative maxima. Barnett's empirical study shows, in fact, that for n between 10 and 20 (n fixed), multiple relative maxima occur in approximately 30 percent of all samples. If multiple solutions do occur infinitely often as $n \to \infty$, then by Theorem 8 all but one solution must approach $\partial\Omega = \{-\infty, \infty\}$.

Example 2: The bivariate normal distribution with unknown correlation but known means and variances.

Let $x = (y, z)$, $\theta = \rho$, Ω = the open interval $(-1, 1)$, and

$$p(x, \rho) = (1-\rho^2)^{-\frac{1}{2}}\exp[-\tfrac{1}{2}(1-\rho^2)^{-1}(y^2-2\rho yz+z^2)] \quad .$$

The LEQ is a cubic equation in ρ (see Kendall and Stuart (1967), p. 38-9) which can have either one or three solutions in Ω. Let ρ_0 denote the true (but unknown) value of the correlation coefficient ρ. Then

$$\frac{\partial \log p(x, \rho)}{\partial\rho} = (1-\rho^2)^{-2}[\rho(1-\rho^2) + yz(1+\rho^2) - \rho(y^2+z^2)]$$

and

$$g_1(\rho) = (1-\rho^2)^{-2}(1+\rho^2)(\rho_0-\rho).$$

The function $g_1(\rho)$ has exactly one zero, at ρ_0, and $g_1(\rho) > 0$ (< 0) if $\rho < \rho_0$ ($> \rho_0$), so this family of densities is strongly K-L ordered. Conditions C1, 4, 5, 6, 8, 9, 10 (but not C10b, c) are satisfied, but C11 does not hold. In this case, however, we can apply the remark at the end of Section 3 with $k(\theta) = k(\rho) = (1-\rho^2)^2$. Then C11 is satisfied when $m(x, \rho)$ is replaced by $m^*(x, \rho)$, and also (3.14) holds since $g_1^*(\rho) = (1+\rho^2)(\rho_0-\rho)$. Therefore the conclusions of Theorem 11 are true, so with probability 1 there is exactly one solution of the LEQ (a relative maximum) in Ω fasln, and this solution converges to ρ_0 as $n \to \infty$. (This proves the strong consistency of the MLE.) This result is obtained by Kendall and Stuart (1967) p. 39, by direct examination of the LEQ.

6. Discussion.

If the LEQ has multiple solutions for some or all n and $(x_1, \cdots, x_n)$, we are faced with the problem of choosing one of these solutions for our estimate of θ_0. If it is known that the MLE is a strongly consistent estimator of θ_0 (as is almost always the case, since Wald's (1949) conditions are quite weak) then we would choose that solution at which the LF achieves its absolute maximum. (The absolute maximum may be a terminal maximum; however, since we assume θ_0 is an interior point of Ω, the absolute maximum will occur at an interior point of Ω fasln.) Thus, two steps are involved -- finding the set $S(x_1, \cdots, x_n)$ of all solutions of the LEQ, and evaluating the LF at each solution. (The usual Newtonian methods, based on successive iterations starting with a preliminary estimate, may converge to a solution other than the absolute maximum or may fail to converge at all -- see Barnett.) Often one or both of these steps are difficult or lengthy to carry out, even with a computer. To insure that all solutions (and hence all maxima) are found, the entire possible range of solutions must be scanned. For example, in the case of the Cauchy distribution, there may exist up to $2n$ solutions, with possible range $[x_{min}, x_{max}]$. Since the Cauchy distribution has heavy "tails" this range may be quite wide, so the procedure of scanning this range for $2n$ possible solutions and evaluating the LF at each may be extremely laborious. (Barnett recommends the method of "false positions," which, although time-consuming, enables us to carry out systematically the scanning and evaluation process and obtain (approximately) the MLE, and which avoids the possibility of failure of convergence. See also Richards (1967).)

If, therefore, we are concerned with a fixed sample size n (moderate) and our objective is to find the MLE, then in general we cannot avoid the necessity of a lengthy scanning procedure if multiple solutions of the LEQ occur. If, however, we are primarily concerned with large sample sizes and are satisfied to obtain an estimator which is asymptotically equivalent to the MLE, then simpler methods are available. For example, it is well-known that if $\tilde{\theta}_n - \theta_0 = O_p(n^{-\frac{1}{2}})$ then the first Newton-Raphson iterate $\theta_n^{(1)}$, starting with $\tilde{\theta}_n$ as a preliminary estimate, is asymptotically equivalent to the MLE $\hat{\theta}_n$ in the (weak) sense that $\theta_n^{(1)} - \hat{\theta}_n = o_p(n^{-\frac{1}{2}})$.

We now propose three methods of a different sort, based on the results of Sections 2 and 3, which are aimed at simplifying the scanning procedure. These methods avoid the necessity of finding all solutions of the LEQ. The first makes use of a preliminary estimator $\bar{\theta}_n$ and only requires that we consider solutions "near" $\bar{\theta}_n$, while the other two methods require only that a single solution be found. (Numerical methods still may be needed to locate a solution.) These three methods provide estimators θ_n^* which are equivalent to the MLE in the very strong sense that $P[\theta_n^* = \hat{\theta}_n \text{ fasln}] = 1$.

First, suppose that the conclusions of Theorem 4 are true and that preliminary information is available in the form of a preliminary estimate $\bar{\theta}_n$ such that $P[\bar{\theta}_n \to \theta_0] = 1$. Nothing need be assumed about the rate of convergence. Let $\theta_n^* = \theta_n^*(x_1, \cdots, x_n)$ be that solution of the LEQ which is closest to $\bar{\theta}_n$. (In practice one would check that θ_n^* is a relative maximum, and if not, one might take the next closest solution; however, this is not necessary.) Then by Theorem 4 and the (assumed) strong consistency of the MLE $\hat{\theta}_n$,

(6.1) $$P[\theta_n^* \text{ is the solution closest to } \theta_0 \text{ fasln}] = 1$$

and

(6.2) $$P[\theta_n^* = \hat{\theta}_n \text{ fasln}] = 1.$$

This method may greatly reduce the range which must be scanned for solutions. If one solution, say θ_n', is found by any method, then we need only scan the interval $[\bar{\theta}_n - d, \bar{\theta}_n + d]$, where $d = |\bar{\theta}_n - \theta_n'|$, to find θ_n^*. Note that the conclusions of Proposition 3 are not strong enough to guarantee the validity of (6.1) and (6.2). This method was suggested to the author by Barnett's empirical study for the Cauchy distribution, where he showed that for small $n (\simeq 3)$ $P[\theta_n^* \neq \hat{\theta}_n] \leq .02$ and for moderate n $(13 \leq n \leq 19)$ $P[\theta_n^* \neq \hat{\theta}_n] \leq .001$ (here, $\bar{\theta}_n$ is the sample median).

If, in addition, the conclusions of Theorem 8 are true, preliminary information of a much less precise nature is sufficient to reduce the range which must be scanned. Suppose it is known only that θ_0 lies in the interior of a compact interval $K_1 \subset \Omega$. Let θ_n^* be any solution of the LEQ lying in K_1, i.e., $\theta_n^*(x_1, \cdots, x_n)$ is chosen arbitrarily from $S(x_1, \cdots, x_n) \cap K_1$. (There may exist no solution in K_1, but with probability 1 there must be a solution in K_1 fasln.) Then (6.1) and (6.2) follow from Theorem 8 and the (assumed) strong consistency of the MLE. This method is applicable, for example, in the case of the Cauchy distribution.

Finally, if the conclusions of Theorems 9, 11, or 12 are true, then <u>no</u> preliminary information is needed. Let θ_n^* be any root of the LEQ in Ω, i.e., $\theta_n^*(x_1, \cdots, x_n)$ is chosen arbitrarily from $S(x_1, \cdots, x_n) \cap \Omega$. In this case we know that the MLE is strongly consistent, so (6.1) and (6.2) again hold. This method is applicable in the second example of Section 5.

The main import of the second and third methods is that if n is large, once we have found a solution (by any means) in K_1 or Ω, respectively, we need not worry that we have found the "wrong" solution. Of course, even if these methods are applicable, the first method might be preferred if a preliminary estimate $\bar{\theta}_n$ is available.

One problem still open is that of studying the distribution of

$N(X_1, \cdots)$, the quantity appearing in Theorems 8, 9, and 11, so we can answer the question "how large must n be to be sufficiently large?"

The methods of this paper can be applied easily to study the limiting behavior of $S(x_1, \cdots, x_n)$ if the family of densities is not strongly K-L ordered, i.e., if $g_1(\theta)$ has several zeroes.

In a future paper we shall treat the multiparameter case where θ is vector-valued.

Acknowledgements. I wish to thank Dr. G. Styan for calling this subject area to my attention and pointing out many of the references cited. Mr. E. Nebenzahl and Professor W. Pruitt made helpful suggestions concerning Lemmas 5 and 10.

REFERENCES

Barnett, V. D. (1966). Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. Biometrika 53, 151-165.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

Hardy, G. H., Littlewood, J. E., and Polya, G. (1952). Inequalities (2nd ed.). Cambridge University Press.

Huzurbazar, V. S. (1948). The likelihood equation, consistency, and the maxima of the likelihood function. Ann. Eugenics 14, 185-200.

Karlin, S. (1968). Total Positivity (Vol. 1). Stanford University Press.

Kendall, M. G., and Stuart, A. (1967). The Advanced Theory of Statistics (Vol. 2, 2nd ed.). Griffin, London.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University of California Publications in Statistics 1, 277-329.

Loeve, M. (1963). Probability Theory (3rd ed.). Van Nostrand, Princeton.

Perlman, M. (1969). The strong consistency of the maximum likelihood estimator. (To appear as a technical report of the Department of Statistics, University of Minnesota.)

Rao, C. R. (1965). Linear Statistical Inference and its Applications. Wiley, New York.

Richards, F. S. G. (1967). On finding local maxima of functions of a real variable. Biometrika 54, 310-311.

Royden, H. L. (1968). Real Analysis (2nd ed.) Macmillan, New York.

Rubin, H. (1956). Uniform convergence of random functions with applications to statistics. Ann. Math. Statist. 27, 200-203.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 595-601.

Appendix

I. We first prove (2.2) of Proposition 2, assuming C4, C5, and C6. (Huzurbazar proved the result under the stronger assumption C6b.) In view of (2.4) it suffices to show that for all $\varepsilon' > 0$,

$$\text{(A.1)} \quad P\left[\left|\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2} - \frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta^2}\right| \le \varepsilon' \text{ fasln}\right] = 1.$$

By the SLLN,

$$\text{(A.2)} \qquad P\left[\frac{1}{n}\sum_{i=1}^{n} M(X_i) \le EM + 1 \text{ fasln}\right] = 1,$$

where $M(x)$ appears in C6. Let $\varepsilon = \varepsilon'(EM + 1)^{-1}$ and let $\delta(\varepsilon) > 0$ be as in C6. Since $\bar{\theta}_n$ is strongly consistent for θ_0

$$P[|\bar{\theta}_n - \theta_0| \le \delta(\varepsilon) \text{ fasln}] = 1,$$

so by C6

$$\text{(A.3)} \quad P\left[\left|\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2}\right] - \frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta^2}\right| \le \frac{\varepsilon}{n}\sum_{i=1}^{n} M(X_i) \text{ fasln}\right] = 1.$$

Combining (A.2) and (A.3) by means of (1.2) we obtain (A.1).

If in Proposition 2 $\bar{\theta}_n$ is only assumed to be a weakly consistent estimator for θ_0, i.e., $\bar{\theta}_n \xrightarrow{p} \theta_0$, then an argument similar to the above (using the Weak LLN) shows that

$$\text{(A.4)} \qquad \frac{1}{n}\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2} \xrightarrow{p} -I(\theta_0).$$

II. Suppose that C4, C5, C6, and (2.1) are true, and that $\{\hat{\theta}_n\}$ is a root of the LEQ which is weakly consistent for θ_0. For any $\varepsilon > 0$ there is an $N(\varepsilon) > 0$ such that $P[\hat{\theta}_n \in I] \ge 1 - \varepsilon$ for all $n \ge N(\varepsilon)$, where I is the neighborhood of θ_0 defined in C4. If $\hat{\theta}_n \in I$, then

(b)

$$0 = \frac{\partial \log L_n(\hat{\theta}_n)}{\partial\theta} = \frac{\partial \log L_n(\theta_0)}{\partial\theta} + (\hat{\theta}_n - \theta_0)\frac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta^2},$$

so

$$\text{(A.5)} \qquad n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = \frac{n^{-\frac{1}{2}}\dfrac{\partial \log L_n(\theta_0)}{\partial\theta}}{n^{-1}\dfrac{\partial^2 \log L_n(\bar{\theta}_n)}{\partial\theta}},$$

where $\bar{\theta}_n = \bar{\theta}_n(x_1, \cdots, x_n)$ lies between $\hat{\theta}_n$ and θ_0. Then $\bar{\theta}_n$ is weakly consistent for θ_0 so by (A.4), the denominator of (A.5) converges in probability to $-I(\theta_0)$. By (2.1) and the Central Limit Theorem the numerator converges in law to the normal distribution $N(0, I(\theta_0))$. Hence

$$L[n^{\frac{1}{2}}(\hat{\theta}_n-\theta_0)] \to N[0, I(\theta_0)^{-1}],$$

so $\hat{\theta}_n$ is asymptotically normal and asymptotically efficient.

III. Here we prove the assertions made in the remark following C10. First we show that C8, C9, and C10 imply that the integral defining $g_1(\theta)$ exists for all θ in Ω. Let $A = \{\theta \in \Omega : g_1(\theta) \text{ exists}\}$ and $B = \Omega - A$. A is nonempty by C9; suppose that B is also nonempty. Since Ω is an interval there must be at least one point $\bar{\theta}$ in Ω which is a boundary point of both A and B. With $\delta = \delta(\varepsilon, \bar{\theta})$ as defined in C10 (any $\varepsilon > 0$ will do), for any points θ_1 and θ_2 in $[\bar{\theta} - \delta, \bar{\theta} + \delta] \cap \Omega$ we have

$$\left|\frac{\partial \log p(x, \theta_1)}{\partial\theta} - \frac{\partial \log p(x, \theta_2)}{\partial\theta}\right| \leq 2\varepsilon\, H_{\bar{\theta}}(x).$$

Since $\bar{\theta}$ is a boundary point of both A and B there must exist two such points θ_1 and θ_2 with $\theta_1 \in A$ and $\theta_2 \in B$. Then, however,

(c)

$$\left|\frac{\partial \log p(x, \theta_2)}{\partial\theta}\right| \le \left|\frac{\partial \log p(x, \theta_1)}{\partial\theta}\right| + 2\epsilon H_{\bar\theta}(x)$$

and the right hand side has finite expectation since $\theta_1 \in A$. Thus $\theta_2 \in A$, a contradiction, so B must be empty and the result is proved.

Under the same assumptions we now show that the K-L information number $I(\theta_0, \theta)$ is finite for all θ in Ω. Recall that $I(\theta_0, \theta) \ge 0$, and let $A = \{\theta \in \Omega: I(\theta_0, \theta) < \infty\}$, $B = \Omega - A$. A is nonempty since $I(\theta_0, \theta_0) = 0$. Suppose that B is also nonempty and let $\bar\theta \in \Omega$ be a boundary point of both A and B. By the Mean Value Theorem, for any θ in Ω

$$\log p(x, \theta) - \log p(x, \bar\theta) = (\theta - \bar\theta)\,\frac{\partial \log p(x, \theta^*)}{\partial\theta},$$

where $\theta^* = \theta^*(x)$ lies between θ and $\bar\theta$. If θ lies in $[\bar\theta - \delta, \bar\theta + \delta] \cap \Omega$ (δ the same as above), then by C10

$$\text{(A.6)} \quad |\log p(x, \theta) - \log p(x, \bar\theta)| \le \delta\left[\left|\frac{\partial \log p(x, \bar\theta)}{\partial\theta}\right| + \epsilon\, H_{\bar\theta}(x)\right]$$

$$\equiv \delta D(x).$$

$D(X)$ has finite expectation by the result of the preceding paragraph. Since $\bar\theta$ is a boundary point of both A and B there must exist two points θ_1 and θ_2, both contained in $[\bar\theta - \delta, \bar\theta + \delta] \cap \Omega$, with θ_1 in A and θ_2 in B. Therefore

$$|\log p(x, \theta_1) - \log p(x, \theta_2)| \le 2\delta D(x),$$

so

$$|\log p(x, \theta_0) - \log p(x, \theta_2)| \le |\log p(x, \theta_0) - \log p(x, \theta_1)| + 2\delta D(x).$$

This last expression has finite expectation since $\theta_1 \in A$, so θ_2 must also be in A, a contradiction. Thus B is empty, proving the result.

Finally, we show that $I(\theta_0, \theta)$ can be differentiated under the integral sign for all θ in Ω, so that $(d/d\theta)I(\theta_0,\theta) = -g_1(\theta)$. For any θ in Ω, C10 implies that there exists $\delta = \delta(\epsilon, \theta) > 0$ (any $\epsilon > 0$ suffices) such that for all θ' in $[\theta - \delta, \theta + \delta] \cap \Omega$,

$$\left|\frac{\partial \log p(x, \theta')}{\partial\theta}\right| \le \left|\frac{\partial \log p(x, \theta)}{\partial\theta}\right| + \epsilon H_\theta(x).$$

Since this expression is integrable and does not depend on θ', the result follows (see Loeve (1963), p. 126, 3^o). A similar argument shows that under C4, C5, and C7, $g_1(\theta)$ can be differentiated under the integral sign in an open interval containing θ_0.

Remark: All of the derivations given in Appendix III remain valid if the null set Λ in C10 is permitted to vary with θ. Furthermore, $g_1(\theta)$ is still continuous on Ω.